1

Introduction to Evidence Synthesis

1.1 Introduction

This chapter gives a broad overview of indirect comparisons and network meta-analysis. Using a worked example, we look at what these methods are for, what they can do, the core assumptions, the kinds of output they can produce and some of the implications for decision-making. These topics are picked up and covered in more detail in later chapters.

We assume that the evidence that has been assembled for meta-analysis has been identified by a protocol-driven systematic review. Although the methods of systematic review are not covered in this book, it will become clear that the underlying assumptions of meta-analysis put certain constraints on the conduct of the systematic review. This will be discussed further, particularly in Chapter 12 on the validity of network meta-analysis. We also assume that only randomised controlled trials (RCTs) will be included in the synthesis, although we include some comments on how observational studies might be included in Chapter 9 on bias models.

We begin by discussing the purpose of indirect comparisons and network meta-analysis, which is to form the basis for coherent, evidence-based treatment decisions. We then look at some simple methods for networks involving three treatments to illustrate the concepts. Next a worked example of a larger network meta-analysis is presented to show the main properties of the method. After mentioning the assumptions made by network meta-analysis, we briefly look at the question of which trials should be included in a network meta-analysis. This turns into a discussion of the participants, intervention, comparators, outcomes (PICO), the standard 'script' that defines inclusion criteria in systematic reviews (Sackett et al., 2000; Higgins and Green, 2008). We briefly consider whether PICO definitions should be reconsidered for the purposes of network meta-analysis.

Network Meta-Analysis for Decision-Making, First Edition. Sofia Dias, A. E. Ades, Nicky J. Welton, Jeroen P. Jansen and Alexander J. Sutton.

Companion website: www.wiley.com/go/dias/networkmeta-analysis

1.2 Why Indirect Comparisons and Network Meta-Analysis?

While pairwise meta-analysis seeks to combine evidence from trials comparing two treatments, A and B, indirect comparisons involve more than two treatments. A variety of structures can be found, some of which are shown in Figure 1.1. More complex structures are also common and will be presented in later chapters. The first use of indirect comparisons was by Bucher et al. (1997a, 1997b). Their objective was to draw inferences about the relative treatment effects of treatments B and C, using data from B versus A (AB) and C versus A (AC) trials. An indirect estimate of d_{BC}, the effect of C relative to B, is formed by subtracting the direct estimate of d_{AB} from the direct estimate of d_{AC}:

$$\hat{d}_{BC}^{Ind} = \hat{d}_{AC}^{Dir} - \hat{d}_{AB}^{Dir} \tag{1.1}$$

While it is true that the indirect comparison provides an estimate of d_{BC}, in practice this may not be the main reason for looking at the AC and AB trials. More often than not, the objective is to consider which is the best of the three treatments. To make such a comparison, based on one AB trial and one AC trial, we have to make the assumption that, although the event rates in the A arm may have been different in the AB and AC trials, the treatment effects, that is, the differences between arms, that are seen in one trial would – subject to sampling error – also be seen in the other. It is for this reason that some

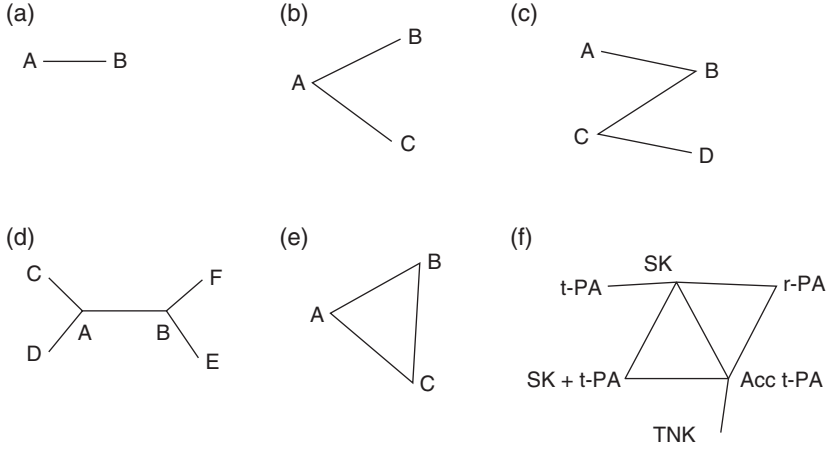

Figure 1.1 Some examples of connected networks: (a) a pairwise comparison, (b) an indirect comparison via reference treatment A, (c) a 'snake' indirect comparison structure, (d) another indirect comparison structure, (e) a simple triangle network and (f) a network of evidence on thrombolytics drugs for acute myocardial infarction (Boland et al., 2003).

authors have referred to this as an 'adjusted' indirect comparison (Glenny et al., 2005), because it takes account of the fact that the AB and AC trials will not necessarily have the same absolute event rate on the A arm. But there is probably no need for this terminology, as the idea of an 'unadjusted' indirect comparison, based on the absolute event rates, has never been seriously considered in the (pairwise) meta-analysis literature as it breaks the fundamental property of randomisation. (A form of unadjusted indirect comparison has, however, been proposed for use in network meta-analysis under the heading 'arm-based models' (Hong et al., 2013, 2015, 2016). These models are reviewed in Chapter 5.)

The assumption that treatment differences on a suitable scale are relatively stable across trials, while absolute event rates can vary, is a fundamental assumption in clinical trials and meta-analysis.

The introduction of 'loops' of evidence into a network represents an additional level of complexity. Examples in Figure 1.1e and f include at least one closed loop of evidence. In the triangle structure of Figure 1.1e, for example, we have two sources of evidence on d_{BC}, direct evidence from the BC trials and indirect evidence from the AB and AC trials. Loops have a special significance as they mean that we are in a position to combine the direct and indirect evidence on d_{BC} and form a single pooled estimate. We are also in a position to assess the degree of consistency (or agreement) between the direct and indirect estimates.

Although we continue to use the terminology 'direct' and 'indirect' throughout the book, these words only have meaning when referring to a specific treatment contrast (we use the term 'contrast' to refer to a pairwise comparison between two treatments). As the network grows in complexity, networks involving over 20 treatments are not uncommon (Corbett et al., 2013; Kriston et al., 2014; Mayo-Wilson et al., 2014), and particularly in a decision-making context, the terms 'direct' and 'indirect' lose all meaning, as evidence that is indirect for one contrast is direct for another, and – depending on network structure – evidence on a single contrast may impact on estimates throughout the whole network. The challenge facing the investigator is first to assemble *all* the trials comparing *any* of the treatments of interest in the target population and second to use the totality of this trial evidence to determine an internally consistent set of estimated relative treatment effects between all treatments. Thus, every trial in the network can be considered as contributing 'direct' evidence towards the decision problem. Furthermore, if we set aside for the moment the results of the trial, its size, and its position in the network structure, we can say that every trial is equally relevant and has an equal chance of influencing the decision.

The primary objective of network meta-analysis is, therefore, simply to combine all the available data in a coherent and internally consistent way. To see what this means, consider the triangle structure of Figure 1.1e. One way

to analyse these data would be to carry out three separate meta-analyses on the AB, AC and BC sets of trials to generate three estimates: $\hat{d}_{AB}^{Dir}, \hat{d}_{AC}^{Dir}, \hat{d}_{BC}^{Dir}$. However, these estimates cannot in themselves form a coherent basis for deciding which treatment is the best. Suppose we had three boys: John, Paul and George. John is 2 inches taller than Paul, and Paul is 3 inches taller than George. From this we know that John is 5 inches taller than George. If we were not able to make this inference, if John could be 6 or 2 inches taller than George, then we would not be able to decide who was tallest.

What is needed, therefore, is not the original, unrelated estimates $\hat{d}_{AB}^{Dir}, \hat{d}_{AC}^{Dir}, \hat{d}_{BC}^{Dir}$, but three coherent estimates $\hat{d}_{AB}^{Coh}, \hat{d}_{AC}^{Coh}, \hat{d}_{BC}^{Coh}$ that have the property: $\hat{d}_{AC}^{Coh} = \hat{d}_{AB}^{Coh} + \hat{d}_{BC}^{Coh}$. With John = A, Paul = B and George = C, we can deduce that the AC difference is 5, the sum of the AB and BC differences. Network meta-analysis is no more than a technical way of processing the trial-specific treatment effects to arrive at the coherent estimates that are required for coherent decisions about which treatment is best while correctly reflecting parameter uncertainty. However, it has to be appreciated that the coherent estimates represent a view of the data that makes some assumptions. These are set out in Section 1.5.

A network meta-analysis can be carried out whenever the included trials and treatments form a connected network, that is, a network where there is a path linking each treatment to every other (Figure 1.1). In that sense, pairwise meta-analysis and indirect comparisons are special cases of simple network meta-analysis. See Chapter 2 for more details on this.

1.3 Some Simple Methods

The first explicit use of indirect comparisons was due to Bucher et al. (1997a, 1997b). An indirect estimate for the treatment effect of C relative to B is formed by subtracting the estimate of B relative to A from the estimate of C relative to A, as in equation (1.1). Before going further, note that we keep to a convention that d_{XY} is the additional effect of Y relative to X and that we always keep treatment indices in numerical or alphabetical order, d_{XY} or d_{35}; $-d_{YX}$ is equivalent to d_{XY}, but we never refer to the former to avoid confusion. A helpful way of remembering how the notation works is to draw the line:

Now consider the relative effect of B compared to A as the 'distance' in relative effect terms between points A and B on the line (hence the notation d_{AB}). This immediately clarifies why the indirect estimate of d_{BC} is calculated as $d_{AC} - d_{AB}$ in equation (1.1), and not $d_{AB} - d_{AC}$. The variance of the indirect estimate is the sum of the variances of the two direct estimates, as

each is derived from different trials including different patients and therefore independent:

$$V_{BC}^{Ind} = V_{AC}^{Dir} + V_{AB}^{Dir} \tag{1.2}$$

See Exercise 1.1. Notice that equation (1.2) tells us that indirect evidence will always have a higher variance than any of its component parts. As the number of links increases, as in Figure 1.1c, the variance of the indirect estimate will become very large (see Exercise 1.3).

We can extend this by pooling the indirect and direct estimates into a single combined estimate. In the triangle structure of Figure 1.1e, we have three sets of trials on AB, AC and BC, and we now conduct three separate pairwise meta-analyses to generate the three pairwise pooled estimates $\hat{d}_{AB}^{Dir}, \hat{d}_{AC}^{Dir}, \hat{d}_{BC}^{Dir}$ and their variances $V_{AB}^{Dir}, V_{AC}^{Dir}, V_{BC}^{Dir}$. We therefore have two independent sources of evidence on d_{BC}, one direct and the other indirect. This invites us to pool the two estimates into a single combined estimate, for example, using inverse-variance weighting:

$$\hat{d}_{BC}^{Pooled} = \frac{w_{BC}^{Dir}\hat{d}_{BC}^{Dir} + w_{BC}^{Ind}\hat{d}_{BC}^{Ind}}{w_{BC}^{Dir} + w_{BC}^{Ind}}, \quad \text{where} \quad w_{BC}^{Dir} = \frac{1}{V_{BC}^{Dir}}$$
$$\text{and} \quad w_{BC}^{Ind} = \frac{1}{V_{BC}^{Ind}} = \frac{1}{V_{AB}^{Dir} + V_{AC}^{Dir}} \tag{1.3}$$

See Exercise 1.2. Before accepting this combined estimate, it will be important to check that the direct and indirect estimates of d_{BC} are consistent with each other. This is explored in Chapter 7.

The aforementioned procedure gives us a pooled estimate of d_{BC}, but it does not immediately yield the set of coherent estimates of all three parameters. We can generate these by simply repeating the process for the other edges. The pooled estimates $\hat{d}_{AB}^{Pooled}$, $\hat{d}_{AC}^{Pooled}$ and $\hat{d}_{BC}^{Pooled}$ will have the required property of coherence, but we can arrive at these estimates more efficiently in a single step using suitable computer programs. Indeed, the simple approach to indirect and mixed treatment comparisons in equations (1.2) and (1.3) is extremely limited: as we go from 3 to 4, 5, 10, 20 or 30 treatments, the number of pairwise comparisons that can be made increases from 3 to 6 and then 15, 45, 190 and finally 445. (For S treatments, there are $S(S-1)/2$ possible contrasts.) It is evident that distinguishing between direct and indirect evidence in the context of a decision on which of several treatments is best is not only a pointless exercise but also a very tedious one. What is needed is a method that can put all this evidence together to produce the coherent estimates required and some methods for assessing the consistency of evidence from different sources.

1.4 An Example of a Network Meta-Analysis

We now develop a worked example of a network meta-analysis. This is taken from a health technology assessment report (Boland et al., 2003) that was published in 2003, based on an analysis of thrombolytic treatments carried out for NICE, but before the use of network meta-analysis had become a routine feature of NICE technology appraisals. The evidence structure (Figure 1.1f) comprises six active treatments: streptokinase (SK), tissue-plasminogen activator (t-PA), accelerated tissue-plasminogen activator (Acc t-PA), SK + t-PA, tenecteplase (TNK) and reteplase (r-PA). Table 1.1 is a logically ordered layout of the evidence structure that makes it easy to see that the majority of the evidence is on trials that compare SK with t-PA. In the original report the authors presented their findings as shown in Table 1.2, which shows the results of four separate pairwise meta-analyses, each undertaken independently, and therefore not having the required property of coherence.

In the absence of a single coherent analysis, the authors were obliged to put forward their conclusions in terms of the pairwise analyses:

> … streptokinase is as effective as non-accelerated alteplase … tenecteplase is as effective as accelerated alteplase … reteplase is at least as effective as streptokinase … [is] streptokinase as effective as, or inferior to accelerated alteplase … [is] reteplase as effective as accelerated alteplase or not … (Boland et al., 2003).

It is difficult to conclude from statements of this sort which treatment performs best. Indeed, given that each estimate is subject to sampling error, under some circumstances, it could prove *impossible* to identify the best treatment using this approach.

Table 1.1 The thrombolytics dataset, 14 trials, six treatments (data from Boland et al., 2003): streptokinase (SK), tissue-plasminogen activator (t-PA), accelerated tissue-plasminogen activator (Acc t-PA), tenecteplase (TNK), and reteplase (r-PA) 14 trials.

RCTs	SK	t-PA	Acc t-PA	SK + t-PA	r-PA	TNK
8	✓	✓				
1	✓		✓	✓		
1	✓			✓		
1	✓				✓	
2			✓		✓	
1			✓			✓

Fifteen possible pairwise comparisons (a tick represents comparisons made in RCTs).
The network is shown in Figure 1.1f.

Table 1.2 Findings from the HTA report on thrombolytics drugs (Boland et al., 2003).

Treatment comparison	Trials	Odds ratio	95% CIs
SK vs t-PA	8	1.00	0.94, 1.06
Acc t-PA vs TNK	1	0.99	0.88, 1.13
Acc t-PA vs r-PA	2	1.24	0.61, 2.53
r-PA vs SK	1	0.94	0.79, 1.12

Table 1.3 Thrombolytics example, fixed effect analysis: odds ratios (posterior medians and 95% CrI).

	SK	t-PA	Acc t-PA	t-PA + SK	r-PA	TNK
SK	–	1.00 (0.94, 1.06)	0.86 (0.78, 0.94)	0.96 (0.87, 1.05)	0.95 (0.79, 1.12)	
t-PA	1.00 (0.94, 1.06)	–				
Acc t-PA	0.86 (0.79, 0.94)	0.87 (0.78, 0.96)	–	1.12 (1.00, 1.25)	1.02 (0.90, 1.16)	1.01 (0.88, 1.14)
t-PA + SK	0.96 (0.88, 1.05)	0.96 (0.86, 1.08)	1.12 (1.00, 1.24)	–		
r-PA	0.90 (0.80, 1.02)	0.91 (0.79, 1.03)	1.04 (0.94, 1.16)	0.94 (0.82, 1.08)	–	
TNK	0.87 (0.75, 1.01)	0.87 (0.74, 1.03)	1.01 (0.89, 1.14)	0.90 (0.77, 1.14)	0.96 (0.82, 1.14)	–

Upper right triangle gives the ORs from pairwise comparisons (column treatment relative to row), lower left triangle ORs from the network meta-analysis (row treatment relative to column).

Table 1.3 sets out the full set of estimated median odds ratio (OR) estimates from a fixed effects network meta-analysis in the lower left triangle and the median ORs from the unrelated pairwise analyses in the upper right triangle, along with their 95% credible intervals. Details on how the network estimates were obtained are given in Chapter 2. The network meta-analysis fills in each of the 15 possible pairwise contrasts. See Exercise 1.4.

It is always informative to look at the relationship between the pairwise estimates and the network estimates. One critical question is whether the direct evidence is consistent with the indirect evidence on each contrast; this cannot be answered from this summary table, but formal methods to detect inconsistency are introduced in Chapter 7. However, there are a number of

useful 'reality' checks that can be carried out. In some parts of the network, for example, the network meta-analysis does not appear to have had much effect: the OR of t-PA relative to SK and its credible interval are entirely unchanged. Examination of the network structure (Figure 1.1f) immediately explains why: the (SK, t-PA) edge is an isolated spur on the network and the SK versus t-PA trials are the sole source of data on this contrast. On the other hand, the network estimate of the (SK, r-PA) comparison has a tighter credible interval than the pairwise estimate, and this is due to the additional indirect evidence contributed by the (SK, Acc t-PA) trials. However, we do not see the same effect with the (SK, SK + t-PA) comparison because this contrast is informed only by a single three-arm trial so that the 'loop' formed by the (SK + t-PA, Acc t-PA) and (SK, Acc t-PA) edges does not constitute additional information. The way that data in one part of the network affects estimates in other parts is determined by the network structure and the quantity of evidence on different edges. Investigators need to be able to 'read' the table in this way to gain an understanding of the main 'drivers' of the results.

The best treatment is Acc t-PA (OR relative to SK 0.86), with TNK a close second (OR 0.87). Following the accepted principles of pairwise meta-analysis, we distinguish between the *relative* treatment effects on mortality, which are pooled in evidence synthesis and are expressed here as ORs, and the absolute mortality on specific treatments. Our view is that absolute effects for given treatments must be considered quite separately, because although trials are essential to inform relative effects, they are not necessarily the best source of information on absolute effects of treatments. This is dealt with in detail in Chapter 5. However, if we can assume a distribution for the absolute 35-day mortality on reference treatment SK, which is appropriate for our target populations, we can use the relative effect estimates from the network meta-analysis to infer the absolute mortality on all the treatments. This is shown in Table 1.4. Given a mortality on SK of 8.33%, we expect 7.29% mortality on Acc t-PA and 7.34% on TNK.

Also shown in Table 1.4 are the posterior median ranks of each treatment (rank '1' indicates lowest mortality) and their credible intervals. The posterior distribution of the rankings and the probability that each treatment is 'best', which is derived from the rankings, are evidently quite unstable. Indeed, the reason for looking at these statistics is only to make oneself aware of the statistical uncertainty in the treatment differences. Decisions should never be based on these probability statements. For example, if a decision was to be made strictly on the grounds of the effect of treatment on a 35-day survival, regardless of cost or side effects, the treatment chosen should be the one that was associated with the lowest expected mortality, in this case Acc t-PA, as shown by the ranking of the posterior *expected* effects in Table 1.4. This is not necessarily the same as the treatment that is most likely to be associated with the lowest mortality, which in this case is TNK (see Exercise 1.5). Similarly,

Table 1.4 Thrombolytics example, fixed effect analysis: posterior summaries.

	Odds ratios		Rank			35 day mortality %		Probability best (%)
	Median	95% CrI	Mean	Median	95% CrI	Mean	95% CrI	
SK	Reference		5.3	5	(4, 6)	8.33	(4.9, 13.1)	0
t-PA	1.00	(0.94, 1.06)	5.1	5	(3, 6)	8.30	(4.9, 13.1)	0.2
Acc t-PA	0.87	(0.79, 0.94)	1.7	2	(1, 3)	7.29	(4.3, 11.6)	41
SK + t-PA	0.96	(0.88, 1.05)	4.1	4	(2, 6)	8.04	(4.7, 12.7)	1.3
r-PA	0.90	(0.80, 1.01)	2.7	3	(1, 5)	7.59	(4.4, 12.1)	16
TNK	0.87	(0.75, 1.01)	2.1	2	(1, 5)	7.34	(4.2, 11.8)	43

Odds ratios, treatment rankings with credible intervals, probability 'best', and absolute mortality. The odds ratios are taken from Table 1.3.

decisions based on net benefit would be based on the treatment with the highest expected net benefit, not on the treatment that is most likely to have the highest net benefit (see Chapter 5). These probability statements can be seen as a way of avoiding the complexity of 15 pairwise significance tests, but should be interpreted with caution. In this case, under the fixed effects model, the data appear to rule out SK, t-PA and SK + t-PA, but do not distinguish between Acc t-PA, TNK, and '3rd place' r-PA.

1.5 Assumptions Made by Indirect Comparisons and Network Meta-Analysis

The assumptions made by network meta-analysis can be stated in several ways. First, in a fixed effects model, one is assuming that the d_{AB} effect that is estimated in the AB trials is the same as the d_{AB} effect that would be estimated in the AC, AD, BC and CD trials *if* these trials had included treatments A and B. With a random effects model, one is assuming that the trial-specific effect $\delta_{i,AB}$, which is estimated in trial i comparing AB, is a sample from a random effects distribution with mean d_{AB} and variance σ^2_{AB}, that all the other AB trials are estimating effects from this same distribution and that if A and B were included in the AC, AD, BC and CD trials, they too would be estimating parameters that are random samples from the same random effects distribution (see Chapter 2 for further details).

Another way to express this is to imagine that all the trials were in fact multiarm trials of treatments A, B, C and D, but only a subset of the treatments are reported and the remaining treatments are missing at random (Little and

Rubin, 2002). This does not mean that every treatment is equally likely to be entered into trials, nor even that missingness is unrelated to efficacy. What is required is that missingness mechanism operates with no regard to the true *relative* efficacy of the treatments (see Chapter 12).

These assumptions are often called the 'consistency assumptions'. Contrary to what is often stated (Song et al., 2009), the consistency assumptions are *not* additional assumptions made by network meta-analysis, but they follow from more basic assumptions underlying all meta-analyses (Lu and Ades, 2009), as explained in Chapter 2 and explored further in Chapter 12. Even so, all the doubts that have been expressed about indirect comparisons and network meta-analysis can be seen as doubts about this assumption. For example:

> Between-trial comparisons [indirect comparisons] are unreliable. Patient populations may differ in their responsiveness to treatment…. Thus an apparently more effective treatment may … have been tested in a more responsive population (Cranney et al., 2002).
>
> Placebo controlled trials lacking an active control give little useful information about comparative effectiveness…. Such information cannot reliably be obtained from cross-study [indirect] comparisons, as the conditions of the studies may have been quite different (ICH Expert Working Group, 2000).

The warnings 'patient populations may differ' or 'the conditions of the study may have been quite different' constitute a general warning that there may be unrecognised effect modifiers, present in the AB trials but not in the AC trials, so that a BC comparison based on this information may be confounded. This is unquestionably correct. But it is not clear that indirect comparisons and network meta-analysis are really different in this aspect from pairwise meta-analysis, where unrecognised effect modifiers are also present, as proven by the frequent need to fit random effects models (Engels et al., 2000), often with extremely high levels of between-study variation (Turner et al., 2012).

Putting this in another way, while there is no doubt that indirect comparisons may be subject to biases arising from failures in internal or external validity, equation (1.1) tells us that indirect evidence (on the left hand side) can only be biased if direct evidence (on the right hand side) is biased. We can unpick this paradox only by reminding ourselves that the concept of 'bias' is relative to a target parameter. For example, suppose our objective is to estimate the effect of C relative to B in patients who are naïve on both B and C, say, $d_{BC}^{(N)}$. We need to be sure that the included AB and AC trials are based on patients who were indeed naïve to both these treatments. If we are satisfied that $\hat{d}_{AB}^{Dir}$ and $\hat{d}_{AC}^{Dir}$ are unbiased estimates of $d_{AB}^{(N)}$ and $d_{AC}^{(N)}$ relating to patients who

are naïve on both B and C, we can be satisfied that $\hat{d}_{BC}^{(N)Ind} = \hat{d}_{AC}^{(N)Dir} - \hat{d}_{AB}^{(N)Dir}$ is also unbiased for this population. If, however, the AC trials are performed on patients who have failed on B, we may think that the direct estimate, $\hat{d}_{AC}^{Dir}$, is 'biased' with respect to the target population, in this case the indirect estimate will inherit this bias.

This, of course, highlights the importance of paying careful attention to which evidence is included in a network meta-analysis (see Chapter 12), or, alternatively, it invites us to consider some form of covariate adjustment (Chapter 8) or bias modelling (Chapter 9). Although we return to this theme in the later chapters, readers may find it useful to see the direction of travel in advance.

The trial data that are available for evidence synthesis are likely, in many instances, to suffer from a range of imperfections. Trials may be only partially *relevant* (external bias) because the patient population is mixed, or the clinical settings are different. Further, the conduct of the trial may have characteristics, such as lack of blinding or allocation concealment, that have been shown to be associated with compromised internal validity (internal bias) (Schulz et al., 1995). This has given rise to a range of methods centred on trial 'quality'. One approach has been the Cochrane Risk of Bias tool (Lundh and Gotzsche, 2008). Another is the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system of assessing the quality of pairwise meta-analyses on a set of pre-determined criteria (Guyatt et al., 2008), extended to deliver quality ratings on network meta-analyses (Puhan et al., 2014; Salanti et al., 2014). These approaches are discussed in more detail in Chapter 12.

Our approach will be somewhat different. Rather than carrying out a network meta-analysis on trial results that we suspect are biased and rating our results from such analysis as unreliable, we would prefer to address any problems with the evidence in advance so as to produce the best possible estimates, whose reliability is appropriately reflected in their credible intervals.

Our strategy to achieve this has several planks. First, knowing that network meta-analysis makes certain assumptions, the most important step is to do everything possible to ensure that the data that are assembled to answer the decision question actually meet those assumptions (see Chapter 12). Second, if there is variation in effect modifiers across the trials in the network, we can use meta-regression (see Chapter 8) to carry out covariate adjustment. Third, if there are potential biases associated with explicit markers of trial 'quality', bias adjustment methods can be applied (see Chapter 9). These can take the form of a regression adjustment, or use can be made of a growing literature that quantifies the extent of potential bias in trials (Wood et al., 2008; Savovic et al., 2012a, 2012b).

1.6 Which Trials to Include in a Network

The existence of methods that can compare multiple treatments, based on all the RCT evidence that compares any of them with any of the others, raises a series of further, more strategic issues. When evaluating whether a new product should be adopted by a health service or whether it should be reimbursed by insurance, one approach is to compare the new therapy to one or more standard therapies and issue a recommendation based on whether it is clinically more effective, or more cost-effective, than standard treatments. Another option is to evaluate the new product against all the standard comparators *and* against all the new competitor products, and then choose the best.

These two procedures appear to be very different, and they clearly have different implications for manufacturers, as the 'choose the best' strategy has the effect of removing all but one of a new set of treatments from the field, and yet both decisions could be based on the same set of trials, and indeed on the same network meta-analysis. This is because the 'evidence space' does not have to coincide with the 'decision space'. The procedure of ranking and choosing the best treatment does not have to be applied to *all* the treatments; it can apply only to the ones that are relevant for the decision, and it is simpler and probably fairer to all concerned, for both decisions to have the same evidential base.

1.6.1 The Need for a Unique Set of Trials

A similar principle lay behind the decision in the methods guide by the National Institute for Health and Clinical Excellence (NICE, 2008b) to permit for the first time the inclusion in evidence networks of treatments that were not of intrinsic interest themselves, but which formed a link between otherwise disconnected networks (Figure 1.2a). If there was no treatment G that had been trialled with at least one of the treatments A, B and C *and* one of the treatments X and Y, then there would be no way of using network meta-analysis to compare X or Y with A, B or C. However, once a treatment G that links the two networks has been identified (Figure 1.2b), are there other treatments that could have made the required connection? If there are, these should also be included, to prevent any suggestions of 'gaming' with the network structure (choosing comparisons that flattered one particular product over the others). The addition of treatment H (Figure 1.2c) illustrates this situation. Finally, if the comparator set is A, B, C, D, G, H, X and Y, we must ask: have all the trials comparing any two or more of these treatments been included? These are no less 'relevant' to the decision problem than the others and should therefore also be included (Figure 1.2d).

To ensure a transparent process that removes any temptation to manipulate trial inclusion/exclusion criteria, a procedure is required, which identifies a

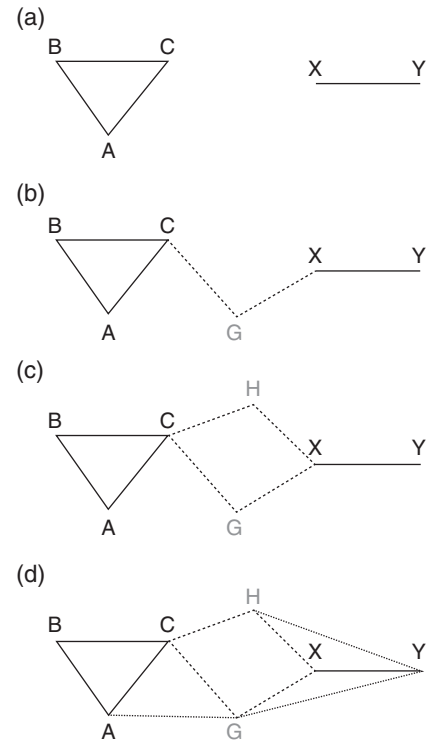

Figure 1.2 (a) An unconnected network; (b) use of an intermediate treatment to link a network (dashed lines); (c) other intermediate treatments that could also be used as links; and (d) incorporation of all trials comparing the enlarged set of treatments.

unique set of trials. Following the illustration in Figure 1.2, we suggest the following algorithm:

1) Identify all the comparators of interest.
2) Identify all the trials that compare two or more of the comparators in the population of interest.
3) Remove trial arms that are not comparators of interest from trials with more than two arms. If this forms a connected network, stop.
4) If this is not a connected network,
   a) Identify further treatments that have been compared with *any* treatments in both sub-networks (including any additional arms that may have been removed in step 3).
   b) Add these to the comparators of interest.
   c) Carry out steps (2) and (3) with the larger set of comparators.

This procedure identifies a minimum set of trials that form a connected network while still being unique. Of course, this does not necessarily guarantee that a connected network can be obtained (see Section 1.7). In addition, a degree of subjective judgement remains as there can always be debate about

the relation between a trial population, often poorly described, and the target 'population of interest', as in any synthesis.

1.7 The Definition of Treatments and Outcomes: Network Connectivity

This raises the question: supposing no connected network of trial evidence can be formed, can one use one-arm studies? Can one use observational studies? Rather than raise false hopes in the reader's mind, we make it clear from the outset that evidence synthesis on treatment effects, as conceived in this book, is based exclusively on randomised trial evidence. Nevertheless, the methods discussed in Chapter 9 on bias models can be applied to observational data on treatment efficacy. These methods force the user to be absolutely explicit about the distribution of potential bias. Similarly, the entire book follows the tradition of meta-analytic methods that concerns the statistical combination of *relative* treatment effects, which rules one-arm studies out of bounds – except when they are used to inform a natural history model. These points are taken up in Chapter 5.

With these provisos, the issue of network connectedness cannot be separated from how treatments and outcomes are defined. In traditional pairwise meta-analysis, this has been determined by the PICO script (Sackett et al., 2000). In the era of network meta-analysis, we suggest in Sections 1.7.1–1.7.3 that there is a need to liberalise the I, C and O constraints on study inclusion while tightening up quite considerably on the P.

1.7.1 Lumping and Splitting

With pairwise meta-analyses it remains common to 'lump' treatments together. Sometimes this is done so there will be enough trials to justify a quantitative synthesis (Gotzsche, 2000). One example of this is a comparison of thrombolytics treatments for acute myocardial infarction against the surgical procedure percutaneous transluminal coronary angioplasty (PTCA) (Keeley et al., 2003b). This analysis, which lumps the thrombolytics treatments (Figure 1.1f), found that PTCA was (statistically) significantly superior to thrombolytics taken as a class. However, in subsequent published correspondence, a 'splitter' objected that this was not the relevant comparison; instead one should compare PTCA with the *best* of the thrombolytics treatments, which was Acc t-PA (Fresco et al., 2003). On this comparison – based on very few trials – PTCA was superior, but this was no longer statistically significant. Revealingly, the original author, evidently a committed 'lumper', declared herself 'mystified' by the criticism of 'lumping trials together' because 'meta-analyses, by definition, pool data' (Keeley et al., 2003a).

In a network meta-analysis one can accept that different thrombolytics may not all be equally effective while at the same time retaining statistical power by combining direct and indirect evidence. A network meta-analysis of the same dataset demonstrated a statistically significant advantage of PTCA over Acc t-PA (Caldwell et al., 2005). See Chapters 2, 3 and 7 for more details on this example.

The default network meta-analysis, therefore, treats every intervention, every dose and every treatment combination a separate 'treatment'. On the other hand, the finer the distinction between treatments, the greater the risk of disconnected networks. There are three ways in which this can be ameliorated: dose response models, treatment combination models or class effects models. The effect of these techniques is to reduce – sometimes dramatically – the number of effective parameters that need to be estimated, and this impacts powerfully on the connectivity of the network. These models are described further in Chapter 8.

1.7.2 Relationships Between Multiple Outcomes

Another way to create more connections is to include more than one 'outcome' in the same synthesis. Trials very frequently report different outcomes at different follow-up times, and some report more than one outcome at more than one time. However, if we restrict ourselves to one particular outcome, or one particular time, we will produce evidence networks that are sparse, poorly connected, or even disconnected. If, on the other hand, the different outcomes can be included in the same analysis, a more connected network is available. What is required is an explicit model for how the treatment effects at different times or on different outcomes are related. However, this is not a trivial undertaking. Chapter 11 gives a number of examples showing how clinical and logical relationships between outcomes can be exploited to provide a single, coherent analysis of all the treatments on all the outcomes and at the same time creating connected networks from unconnected components.

1.7.3 How Large Should a Network Be?

There is no theoretical limit to how large a network can be, and a question that is often asked is 'how large *should* a network be?' In theory, the larger the network, the more robust the conclusions will be, in the sense that the conclusions will become increasingly insensitive to the inclusion or exclusion of any particular trial. However, there is little doubt that the larger the network becomes, the greater the risk of clinical heterogeneity in the trial populations, particularly as one reaches further back to include older trials. This is because new treatments tend to be introduced for more severely ill patients. As time goes on more data on side effects accumulates, and eventually clinicians start to feel it might benefit less ill patients who are then included in later trials.

For this reason it may often turn out that larger networks will fail to produce more precise estimates: greater clinical heterogeneity may lead to greater statistical heterogeneity in treatment effects, larger between-trial variances in random effects models and therefore lower precision in estimates of mean treatment effects (Cooper et al., 2011). It may also increase the risk of inconsistency, as older treatments will have been compared on more severely ill populations than newer treatments. Thus different sources of evidence (direct and indirect) may conflict.

1.8 Summary

In this introductory chapter we have looked at why indirect comparisons and network meta-analysis are needed. We have emphasised that their key property is that they produce an internally coherent set of estimates, which is essential for rational decision-making, whether based on efficacy of cost-effectiveness. We have observed that the simpler methods, which can be applied to triangular networks, have limited scope, and they become completely unfeasible as the networks grow in complexity. We have set out results from a worked example and shown how to interpret them and how comparison with pairwise meta-analysis results can help us understand the main 'drivers' of the results. Finally we have set out the key assumptions in an informal way and looked at some of the strategic policy issues that the existence of network meta-analysis immediately raises: what treatments should be compared, which trials should be included, and how large should a network of comparisons be.

As we work through the book, many of these themes will be picked up again and covered in more detail.

1.9 Exercises

1.1 This table gives results, in the form of log hazard ratios from a study of virologic suppression following highly active antiretroviral therapy for HIV (Chou et al., 2006). Obtain an indirect estimate for $\hat{d}_{BC}$.

	Log hazard ratio	95% CI	Standard error
$\hat{d}_{BC}$	0.47	0.27, 0.67	0.10
$\hat{d}_{AB}$	2.79	1.69, 3.89	0.56
$\hat{d}_{AC}$	1.42	–0.76, 2.08	0.34

1.2 Pool the direct and indirect estimates on BC from the previous exercise. Comment on the validity of the pooled estimate.

1.3 Find an indirect estimate for $\hat{d}_{AD}$ and its standard error using the data below. (Hint: draw the treatment network first.)

	Difference in median survival	Standard error
$\hat{d}_{AB}$	−2.8	1.42
$\hat{d}_{BC}$	2.7	1.24
$\hat{d}_{CD}$	3.0	1.20

1.4 The table below shows posterior mean and median odds ratio from two sets of analyses on a triangle network of three treatments A, B and C. One of these analyses represents pairwise summaries from unrelated meta-analyses, while the other comes from a network meta-analysis. Which is which?

	Analysis 1		Analysis 2	
	Mean	Median	Mean	Median
OR_{AB}	1.773	1.624	1.669	1.405
OR_{AC}	2.378	2.289	2.444	2.341
OR_{BC}	1.540	1.411	1.249	0.9538

1.5 *In Section 1.4 it was noted that the treatment with the highest expected treatment effect may not always be the same as the treatment that is most likely to have the highest treatment effect. Under what circumstances can this happen?